

Для пояснення людських дій в якості стрижневого доцільно розглядати запитання: «У відповідності з якими підставами Ви діяли як раціональна особистість?». Відповідь на це питання містить раціональне пояснення того, на яких підставах діяла раціональна особистість в умовах розриву, і має логічну форму: « $X$  здійснив дію  $A$  на підставі  $R$ », або, в розгорнутому вигляді:

Нехай існує  $x$ , де  $x$  = особистість  $S$ ; існує  $y$ , де  $y$  = дія  $A$ ; існує  $z$ , де  $z$  = підстава  $R$ .  $X$  здійснив  $y$ , причому, здійснюючи  $y$ ,  $x$  діяв у відповідності до  $z$  [1, с. 107–109].

Таким чином, раціональне розуміння та пояснення дії зводиться до встановлення її підстав, в якості яких можуть виступати елементи більш ранньої поведінки як самого індивіда, так й інших індивідів, зокрема їх бажання, віра, зобов'язання, інтенції тощо.

### Список використаних джерел

1. Серль Дж. Рациональность в действии ; пер. в англ. А. Колодия, Е. Румянцевой / Дж. Серль. – М. : Прогресс-Традиция, 2004. – 336 с.

## **МОДЕЛЮВАННЯ ПРОЦЕСІВ ПОШУКУ ТА КЛАСИФІКАЦІЇ СЛУЖБОВИХ ДОКУМЕНТІВ В АВТОМАТИЗОВАНИХ ІНФОРМАЦІЙНО-ПОШУКОВИХ СИСТЕМАХ**

*Л. М. Колєчкіна, д. ф.-м. н., професор;*

*А. І. Бехтольд, магістр*

*ВНЗ Укоопспілки «Полтавський університет економіки і торгівлі»  
м. Полтава, Україна*

На сьогоднішній день розроблено достатньо спеціальних програмно-апаратних засобів, які беруть на себе основні аспекти роботи по зберіганню, обробці, пересиланню документів [1–6]. Отримання та пошук інформації відіграє важливу роль у широкому діапазоні задач управління інформацією та задач електронної комерції. Перспективам розробки інтерактивних пошукових систем присвячена робота [5], у якій зазначено, що не зважаючи на важливість отримання інформації, інформаційні пошукові системи часто слабо відрізняються від перспектив взаємодії людини з комп'ютером. Саме тому необхідно оцінювати ефективність систем електронного документообігу за різними критеріями [2]. Однією з проблем, які постають перед автоматизованими систе-

мами описаного типу є проблема машинного розуміння природної мови. У сучасних документальних інформаційно-пошукових системах відповідність між запитом користувача та документом виконується комп'ютером, що практично виключає використання природної мови у якості основного засобу представлення інформації. Це пояснюється недоліками природної мови з точки зору машинної технології обробки інформації [3].

Перспективним інструментом розв'язання проблем у цьому напрямі є семантичний аналіз, який знаходить своє застосування у різних галузях [4, 5]. Під цим терміном мається на увазі підходи, методи, моделі комп'ютерної обробки текстів з метою виявлення значення цього тексту, виявлення тематики, пошуку схожих текстів.

Трудомісткість операції класифікації одного документу складається з витрат на:

1. Обчислення оцінки близькості документа до даної тематики (для кожної тематики):

- Пошук необхідної інформації про кожен терм з опису тематики і документа.

- Обчислення оцінки близькості двох термів.

- Обчислення загальної оцінки близькості документу тематиці.

2. Вибору найбільш близької тематики.

Загальна трудомісткість класифікації одного документу складає порядку  $O(|\Omega||D|_{avg}|C|W|k)$  операцій, де  $|\Omega|$  – загальне число тематик,  $D_{avg}$  – середня кількість термів в документі,  $|C|$  – середня кількість термів в описі тематики,  $|W|$  – число термів в загальному словнику,  $k$  – розмірність простору гіпотез (число використовуваних сингулярних значень матриці терми-на-документи). Таким чином хоча описаний підхід вимагає значних обчислювальних ресурсів на підготовчому етапі, обчислювальна трудомісткість етапу класифікації відносно невелика.

Для подальшого поліпшення якості класифікації як правило, досліджується ряд ідей.

1. Багаторівнева класифікація. Багато труднощів при класифікації викликано тим фактом, що деякі тематики досить близькі одна до одної, що погіршує результати класифікації. Для вирішення цієї проблеми пропонується використовувати багаторівневий підхід [5]:

- виявлені групи дуже близьких тематик об'єднуються в мегатематики;

- проводиться класифікація по отриманій множині мегатематик;

- для кожної мегатематики проводиться додаткова класифікація документів, що потрапили в неї. Встановлено, що такий підхід дозволяє значно підвищити точність класифікації на етапі процесу класифікації по мегатематикам.

2. Облік зворотного зв'язку. Перспективним методом поліпшення якості класифікації є облік коментарів користувачів системи для точнішого обчислення оцінок тематичної близькості. Такий підхід називається механізмом зворотного зв'язку (relevance feedback) [6].

3. Вибір документів для завдання тематики. Набір документів, використовуваних для завдання тематики, в значній мірі визначає набір слів, які використовуватимуться як опис даної тематики, а також побічно впливає на описи інших тематик. Весь набір документів, використовуваних для завдання тематик, також визначає загальний словник і функцію тематичної близькості. Під час досліджень, що проводилися під час написання цієї роботи була виявлена можливість застосування моделі Google Page Rank також і для обробки документів на митну тематику. Як відомо, документи мають чітку структуру. Крім того, документи митного спрямування не з'являються та не існують ізольовано від інших. Вони можуть бути створені на основі вже існуючого документу, можуть його доповнювати, модифікувати, припиняти дію.

Отже, можна виділити такі відношення між документами:

- створений на основі;
- відношення модифікації;
- скасування дії документу.

Аналогічно, вище зазначені відношення можна використовувати для аналізу взаємозв'язків між документами. Виявлену властивість також є доцільним використовувати для ранжування документів при пошуку, побудови ієрархії документів, класифікації або кластеризації документів. Крім того, відношення «скасування дії» та «модифікації» можуть використовуватися для актуалізації бази даних документів, тобто підтримання їх у стані, що відповідає чинному законодавству.

У відповідності до мети роботи, досліджено застосування методу латентно-семантичного аналізу до пошуку та класифікації документів; виявлена й обґрунтована можливість застосування

методу аналізу взаємозв'язків Google Page Rank [3, 6]. Є доцільним вироблення рекомендацій щодо спільного застосування латентно-семантичного аналізу та Google Page Rank до пошуку документів на обрану тематику. Проаналізовані вище підходи є перспективними для подальших теоретичних та практичних досліджень. Зокрема, перспективами можна вважати: програмну реалізацію методу латентно-семантичного аналізу для побудови документальної пошуково-інформаційної системи документів; побудова за допомогою Google PageRank ієрархічної моделі взаємозв'язків між нормативно-правовими актами, що характеризують їх важливість в даній тематиці; побудова експертами структури тематик, створення словників для них і класифікація документів по заданим тематикам.

### Список використаних джерел

1. Ульяновська Ю. В. Автоматизація діловодства в митній справі [Текст] / Ю. В. Ульяновська, В. О. Яковенко, В. М. Ганжа // Вісник Академії митної служби України. – 2006. – № 1 (29). – С. 77–80.
2. Величкевич М. Б. Електронний документообіг, тенденції та перспективи [Текст] / М. Б. Величкевич, Н. В. Мітрофан, Н. Е. Кунанець // Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі. – 2010. – № 689. – С. 44–54.
3. Матвієнко О. В. Основи організації електронного документообігу [Текст] / О. В. Матвієнко, М. Н. Цивін. – К. : Центр учбової літератури, 2008. – 112 с.
4. Корнеев В. В. Базы данных. Интеллектуальная обработка информации [Текст] / В. В. Корнеев, А. Ф. Гареев, С. В. Васютин, В. В. Райх. – М. : «Нолидж», 2000. – 352 с.
5. Круковский М. Ю. Критерии эффективности систем электронного документооборота [Текст] / М. Ю. Круковский // Системи підтримки прийняття рішень. Теорія і практика, 2005. – С. 107–111.
6. Кураленок И. Автоматическая классификация документов на основе латентно-семантического анализа [Текст] / И. Кураленок, И. Некрестьянов // Научные труды Донецкого национального технического университета. Серия: Информатика, кибернетика и вычислительная техника (ИКВТ-2006). – Вып. 25. – Донецк : ДонНТУ, 2006. – С. 324–335.